

Overview

HP supports, on select HP ProLiant servers, computational accelerator modules based on NVIDIA® Tesla™ Graphical Processing Unit (GPU) technology.

The following Tesla Modules are available from HP, for use in certain HP SL-series servers.

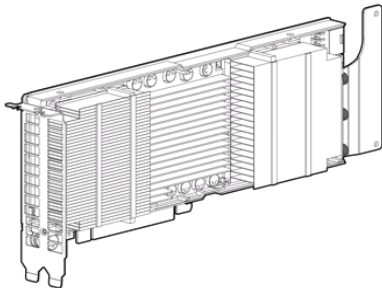
- NVIDIA Tesla M2070Q 2-Slot Passive Module
- NVIDIA Tesla M2090 2-Slot Passive Module
- NVIDIA Tesla K10 Dual GPU PCIe Module
- NVIDIA Tesla K10 Rev B Dual GPU Module
- NVIDIA Tesla K20 5 GB Module
- NVIDIA Tesla K20X 6 GB Module

The NVIDIA Tesla M2070Q module can also be used in HP ProLiant WS460c workstation blades.

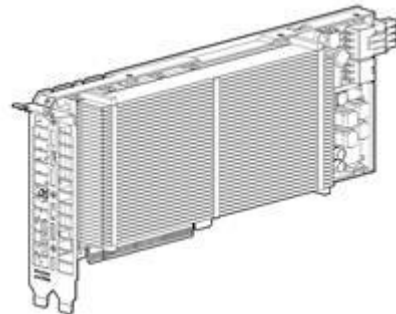
Based on NVIDIA's CUDA™ architecture, the Tesla Modules enable seamless integration of GPU computing with HP ProLiant servers for high-performance computing and large data center, scale-out deployments. These Tesla Modules deliver all of the standard benefits of GPU computing while enabling maximum reliability and tight integration with system monitoring and management tools such as HP Insight Cluster Management Utility.

The Tesla M2070Q uses the NVIDIA Fermi GPU architecture that combines Tesla's high performance computing - found in the other Tesla Modules - and the NVIDIA Quadro® professional-class visualization in the same GPU. The Tesla M2070Q is the ideal solution for customers who want to deploy high performance computing in addition to advanced and remote visualization in the same datacenter.

The HP GPU Ecosystem includes HP Cluster Platform specification and qualification, HP-supported GPU-aware cluster software, and also third-party GPU-aware cluster software for NVIDIA Tesla Modules on HP ProLiant Servers. In particular, the HP Insight Cluster Management Utility (CMU) will monitor and display GPU health sensors such as temperature. CMU will also install and provision the GPU drivers and the CUDA software. The HP HPC Linux Value Pack includes a GPU-enhanced version of IBM Platform LSF, with the capability of scheduling jobs based on GPU requirements. This capability is also available for HP in other popular schedulers such as Altair PBS Professional, and Adaptive Moab.



NVIDIA K10



NVIDIA K20, K20X

What's New

- Support for the NVIDIA Tesla K10 Rev B Dual GPU Module

Models

NVIDIA Passive Tesla Modules

NVIDIA Tesla M2070Q 6GB GPU Graphics Module	A0C39A
NVIDIA Tesla M2075 6 GB Computational Accelerator	A0R41A
NVIDIA Tesla M2090 6 GB Computational Accelerator	A0J99A
NOTE: 2-slot passively cooled Tesla modules with 6 GB memory. NOTE: Please see the HP ProLiant SL250s Gen8 or SL270s Gen8 server QuickSpecs for Technical Specifications and additional information. http://h18004.www1.hp.com/products/quickspecs/14232_na/14232_na.html http://h18004.www1.hp.com/products/quickspecs/14405_na/14405_na.html	
NVIDIA Tesla K10 Dual GPU PCIe Computational Accelerator	B3M66A
NOTE: Please see the HP ProLiant SL250s Gen8 or SL270s Gen8 QuickSpecs for Technical Specifications and additional information. http://h18004.www1.hp.com/products/quickspecs/14232_na/14232_na.html http://h18004.www1.hp.com/products/quickspecs/14405_na/14405_na.html	
NVIDIA Tesla K10 Rev B Dual GPU PCIe Computational Accelerator	E5V47A
NOTE: 2-slot passively cooled pair of Tesla GPUs each with 4 GB memory. NOTE: Please see the HP ProLiant SL250s Gen8 server QuickSpecs for Technical Specifications and additional information. http://h18004.www1.hp.com/products/quickspecs/14232_na/14232_na.html http://h18004.www1.hp.com/products/quickspecs/14405_na/14405_na.html	
NVIDIA Tesla K20 5 GB Computational Accelerator	C7S14A
NVIDIA Tesla K20X 6 GB Computational Accelerator	C7S15A
NOTE: 2-slot passively cooled Tesla GPUs based on NVIDIA Kepler architecture. NOTE: Please see the HP ProLiant SL250s Gen8 server QuickSpecs for Technical Specifications and additional information. http://h18004.www1.hp.com/products/quickspecs/14232_na/14232_na.html http://h18004.www1.hp.com/products/quickspecs/14405_na/14405_na.html	

Standard Features

M2070Q, M2090, K10, K20 and K20X Modules

Performance of the M2090 Module

- 512 CUDA cores
- 655 Gigaflops of double-precision peak performance
- 1330 Gigaflops of single-precision peak performance
- GDDR5 memory optimizes performance and reduces data transfers by keeping large data sets in 6 GB of local memory that is attached directly to the GPU.
- The NVIDIA Parallel DataCache™ accelerates algorithms such as physics solvers, ray-tracing, and sparse matrix multiplication where data addresses are not known beforehand. This includes a configurable L1 cache per Streaming Multiprocessor block and a unified L2 cache for all of the processor cores.
- The NVIDIA GigaThread™ Engine maximizes the throughput by faster context switching that is 10X faster than the M1060 module, concurrent kernel execution, and improved thread block scheduling.
- Asynchronous transfer turbo charges system performance by transferring data over the PCIe bus while the computing cores are crunching other data. Even applications with heavy data-transfer requirements, such as seismic processing, can maximize the computing efficiency by transferring data to local memory before it is needed.
- The high speed PCIe Gen 2.0 data transfer maximizes bandwidth between the HP ProLiant server and the Tesla processors.

Performance of the K10 Module

- 3072 CUDA cores (1536 per GPU)
- 190 Gigaflops of double-precision peak performance (95 Gflops in each GPU)
- 4577 Gigaflops of single-precision peak performance (2288 Gigaflops in each GPU)
- GDDR5 memory optimizes performance and reduces data transfers by keeping large data sets in 8 GB of local memory, 4 GB attached directly to each GPU.
- The NVIDIA Parallel DataCache™ accelerates algorithms such as physics solvers, ray-tracing, and sparse matrix multiplication where data addresses are not known beforehand. This includes a configurable L1 cache per Streaming Multiprocessor block and a unified L2 cache for all of the processor cores.
- Asynchronous transfer turbo charges system performance by transferring data over the PCIe bus while the computing cores are crunching other data. Even applications with heavy data-transfer requirements, such as seismic processing, can maximize the computing efficiency by transferring data to local memory before it is needed.
- The high speed PCIe Gen 3.0 data transfer maximizes bandwidth between the HP ProLiant server and the Tesla processors.

Performance of the K20 Module

- 2496 CUDA cores
- 1.17 Tflops of double-precision peak performance
- 3.52 Tflops of single-precision peak performance
- GDDR5 memory optimizes performance and reduces data transfers by keeping large data sets in 5 GB of local memory that is attached to the GPU
- The NVIDIA Parallel DataCache™ accelerates algorithms such as physics solvers, ray-tracing, and sparse matrix multiplication where data addresses are not known beforehand. This includes a configurable L1 cache per Streaming Multiprocessor block and a unified L2 cache for all of the processor cores.
- Asynchronous transfer turbo charges system performance by transferring data over the PCIe bus while the computing cores are crunching other data. Even applications with heavy data-transfer requirements, such as seismic processing, can maximize the computing efficiency by transferring data to local memory before it is needed.
- Dynamic Parallelism capability that enables GPU threads to automatically spawn new threads.
- Hyper-Q feature that enables multiple CPU cores to simultaneously utilize the CUDA cores on a single GPU.
- The high speed PCIe Gen 2.0 data transfer maximizes bandwidth between the HP ProLiant server and the Tesla processors.

Standard Features

Performance of the K20X Module

- 2688 CUDA cores
- 1.32 Tflops of double-precision peak performance
- 3.95 Tflops of single-precision peak performance
- GDDR5 memory optimizes performance and reduces data transfers by keeping large data sets in 6 GB of local memory that is attached to the GPU
- The NVIDIA Parallel DataCache™ accelerates algorithms such as physics solvers, ray-tracing, and sparse matrix multiplication where data addresses are not known beforehand. This includes a configurable L1 cache per Streaming Multiprocessor block and a unified L2 cache for all of the processor cores.
- Asynchronous transfer turbo charges system performance by transferring data over the PCIe bus while the computing cores are crunching other data. Even applications with heavy data-transfer requirements, such as seismic processing, can maximize the computing efficiency by transferring data to local memory before it is needed.
- Dynamic Parallelism capability that enables GPU threads to automatically spawn new threads.
- Hyper-Q feature that enables multiple CPU cores to simultaneously utilize the CUDA cores on a single GPU.
- The high speed PCIe Gen 2.0 data transfer maximizes bandwidth between the HP ProLiant server and the Tesla processors.

Reliability

- "ECC Memory meets a critical requirement for computing accuracy and reliability for datacenters and supercomputing centers. It offers protection of data in memory to enhance data integrity and reliability for applications. For M2075, M2070Q, M2090, K20, K20X register files, L1/L2 caches, shared memory, and DRAM all are ECC protected. For K10, only external DRAM is ECC protected. Double-bit errors are detected and can trigger alerts with the HP Cluster Management Utility. Also, the Platform LSF job scheduler, available as part of HP HPC Linux Value Pack, can be configured to report when jobs encounter double-bit errors.
- Passive heatsink design eliminates moving parts and cables reduces mean time between failures.

Programming and Management Ecosystem

- The CUDA programming environment has broad support of programming languages and APIs. Choose C, C++, OpenCL, DirectCompute, or Fortran to express application parallelism and take advantage of the innovative Tesla architectures. The CUDA software, as well as the GPU drivers, can be automatically installed on HP ProLiant servers, by HP Insight Cluster Management Utility.
- "Exclusive mode" enables application-exclusive access to a particular GPU. CUDA environment variables enable cluster management software such as the Platform LSF job scheduler (available as part of HP HPC Linux Value Pack) to limit the Tesla GPUs an application can use.
- With HP ProLiant servers, application programmers can control the mapping between processes running on individual cores, and the GPUs with which those processes communicate. By judicious mappings, the GPU bandwidth, and thus overall performance, can be optimized. The technique is described in a white paper available to HP customers at: www.hp.com/go/hpc. A heuristic version of this affinity-mapping has also been implemented by HP as an option to the mpirun command as used for example with HP-MPI, available as part of HP HPC Linux Value Pack.
- GPU control is available through the nvidia-smi tool which lets you control compute-mode (e.g. exclusive), enable/disable/report ECC and check/reset double-bit error count. IPMI and iLO gather data such as GPU temperature. HP Cluster Management Utility has incorporated these sensors into its monitoring features so that cluster-wide GPU data can be presented in real time, can be stored for historical analysis and can be easily used to set up management alerts.

Standard Features

Supported Operating Systems

NOTE: The NVIDIA Tesla modules are supported only on 64-bit versions of Linux and Windows operating systems. The supported operating systems are those below.

RHEL 5

RHEL 6

SLES 11

Windows Server 2008

Supported Servers and Workstation Blades

HP ProLiant SL250s, (K10, K20, K20X)

NOTE: The ambient temperature for SL250s systems with between one and three NVIDIA must be 30 degrees Celsius or less.

NOTE: Consult an HP Solution Architect for precise configuration rules.

HP ProLiant SL270s (K10, K20, K20X)

NOTE: The ambient temperature for SL270s systems with between five and eight NVIDIA must be 30 degrees Celsius or less. All other SL270s systems may be operated with ambient temperatures up to 35 degrees Celsius

NOTE: Consult an HP Solution Architect for precise configuration rules.

HP Services and Support

The NVIDIA Tesla GPU Module has one year for parts exchange only or the warranty of the server or chassis it is attached to and for which it is qualified.

Enhancements to warranty services are available for server and chassis through Flexible Care Pack services.

NOTE: For more information, visit HP Care Pack Services at: <http://www.hp.com/services>

Optional Features

HP High Performance Clusters	HP Cluster Platforms	HP Cluster Platforms are specifically engineered, factory-integrated large-scale ProLiant clusters optimized for High Performance Computing, with a choice of servers, networks and software. Operating system options include specially priced offerings for Red Hat Enterprise Linux and SUSE Linux Enterprise Server, as well as Microsoft Windows HPC Server. A Cluster Platform Configurator simplifies ordering. http://www.hp.com/go/clusters
	HP HPC Interconnects	High Performance Computing (HPC) interconnect technologies are available for this server as part of the HP Cluster Platform portfolio. These high-speed InfiniBand and Gigabit interconnects are fully supported by HP when integrated within an HP cluster. Flexible, validated solutions can be defined with the help of configuration tools. http://www.hp.com/techservers/clusters/ucp/index.html
	HP Insight Cluster Management Utility	HP Insight Cluster Management Utility (CMU) is an HP-licensed and HP-supported suite of tools that are used for lifecycle management of hyperscale clusters of Linux ProLiant systems. CMU includes software for the centralized provisioning, management and monitoring of nodes. CMU makes the administration of clusters user friendly, efficient, and effective. http://www.hp.com/go/icmu
	HP HPC Linux Value Pack	HP HPC Linux Value Pack (Value Pack) is an HP-licensed and HP-supported option to HP Insight CMU, for the development and deployment of applications on HPC Cluster Platforms. Value Pack includes the Platform LSF workload scheduler and Platform Application Center, as well as the HP-MPI parallelization library. HP HPC Linux Value Pack

Third Party GPU Cluster and Development Software

More software for applications and development tools for general purpose GPU enabled systems are available every week. Examples of software available for various vendors are listed below.

- PGI Accelerator: Fortran and C Compilers (directive-based generation of CUDA code, and additionally a CUDA Fortran compiler)
- CAPS HMPP C and Fortran to CUDA C Compiler (directive-based generation of CUDA code)
- TotalView Dynamic Source Code and Memory Debugging for C, C++ and FORTRAN HPC Applications
- Allinea DDT Distributed Debugging Tool
- Wolfram Mathematica mathematical analysis software
- Altair PBS Professional workload scheduler
- Platform LSF workload scheduler
- Adaptive Computing Moab scheduler
- Microsoft Windows HPC Server 2008

Service and Support

HP Technology Services

HP Technology Services offers you technical consultants and support expertise to solve your most complex infrastructure problems. We help keep your business running, boost availability, and avoid downtime.

Protect your business beyond warranty with HP Care Pack Services

When you buy HP Options, it's also a good time to think about what level of service you may need. HP Care Pack services provide total care and support expertise with committed response choices designed to meet your IT and business need.

Optional Features

HP Foundation Care services deliver scalable support-packages for HP industry-standard servers and software. You can choose the type and level of service that is most suitable for your business needs. New to this portfolio is HP Collaborative Support. If you are running business critical environments, HP offers Proactive Care or Critical Advantage. These services help you deliver high levels of application availability through proactive service management and advanced technical response.

Here is the support service recommendation from the Foundation Care and Proactive Care portfolio. For customized support service solution, HP can work with you to tailor a service solution for your unique support requirements using broader services portfolio of Foundation Care and Proactive Care.

Recommended HP Care Pack Services for optimal satisfaction with your HP product

Recommended Services **3-Year HP 24x7 4 hour Response, Proactive Care**

Helps optimize your systems and delivers high levels of application availability through proactive service management and advanced technical response. A skilled Technical Manager will own your query or issue end to end until resolved, delivering a single point of contact for you.

OR

3-Year HP 24x7 4 hour Response, HP Collaborative Support

Provides problem resolution support across the stack of HW, firmware, and HP and 3rd party SW. In case the issue is with 3rd party SW, HP does known issue resolution. If HP cannot solve the issue, it will contact the third-party vendor and create a problem incident on your behalf
<http://h20195.www2.hp.com/V2/GetPDF.aspx/4AA3-8232ENW.pdf>

HP Installation of ProLiant Add On Options Service

This easy-to-buy, easy-to-use HP Care Pack service helps ensure that your new HP hardware or software is installed smoothly, efficiently, and with minimal disruption of your IT and business

Related HP Care Pack Services to enhance your HP product experience

Related Services **3-Year HP 24x7 4 hour Response, Proactive Care**

Helps optimize your systems and delivers high levels of application availability through proactive service management and advanced technical response. A skilled Technical Manager will own your query or issue end to end until resolved, delivering a single point of contact for you

3-Year HP 24x7 4 hour Response, Hardware Support Onsite Service

Provides you with rapid remote support and if required an HP authorized representative who will arrive on site any time and day of the year to begin hardware maintenance service within 4 hours of the service request being logged
<http://h20195.www2.hp.com/V2/GetPDF.aspx/5982-6547EEE.pdf>

3-Year HP 6-hour Onsite Call-to-Repair, HP Collaborative Support

Offers customers a single point of contact for server problem diagnosis, hardware problem resolution to return the hardware in operating condition within 6 hours of the initial service request to the HP Global Solution Center, and basic software problem diagnosis, fault isolation, and resolution if available to HP.
<http://h20195.www2.hp.com/V2/GetPDF.aspx/4AA3-8232ENW.pdf>

HP Proactive Select Service

Provides a flexible way to purchase HP best-in-class consultancy and technical services. You can buy Proactive Select Service Credits when you purchase your hardware and then use the credits over the next 12 months. <http://h20195.www2.hp.com/V2/GetPDF.aspx/4AA2-3842ENN.pdf>

Optional Features

Insight Remote Support Provides 24 X 7 remote monitoring, proactive notifications, and problem resolution. Learn more <http://www.hp.com/go/insightremotesupport>

HP Support Center Personalized online support portal with access to information, tools and experts to support HP business products. Submit support cases online, chat with HP experts, access support resources or collaborate with peers. Learn more <http://www.hp.com/go/hpsc>

HP's Support Center Mobile App* allows you to resolve issues yourself or quickly connect to an agent for live support. Now, you can get access to personalized IT support anywhere, anytime.

HP Insight Remote Support and HP Support Center are available at no additional cost with a HP warranty, HP Care Pack or HP contractual support agreement.

*HP' Support Center Mobile App is subject to local availability

Parts and materials HP will provide HP-supported replacement parts and materials necessary to maintain the covered hardware product in operating condition, including parts and materials for available and recommended engineering improvements. Supplies and consumable parts will not be provided as part of this service; standard warranty terms and conditions apply. Parts and components that have exceeded their maximum supported lifetime and/or the maximum usage limitations as set forth in the manufacturer's operating manual or the technical product data sheet will not be provided, repaired or replaced as part of this service

Warranty / Service Coverage For ProLiant servers and storage systems, this service covers HP-branded hardware options qualified for the server, purchased at the same time or afterward, internal to the enclosure, as well as external monitors up to 22" and tower UPS products; these items will be covered at the same service level and for the same coverage period as the server unless the maximum supported lifetime and/or the maximum usage limitation has been exceeded. Coverage of the UPS battery is not included; standard warranty terms and conditions apply.

The defective media retention service feature option applies only to Disk or eligible SSD/Flash Drives replaced by HP due to malfunction. It does not apply to any exchange of Disk or SSD/Flash Drives that have not failed. SSD/Flash Drives that are specified by HP as consumable parts and/or that have exceeded maximum supported lifetime and/or the maximum usage limit as set forth in the manufacturer's operating manual or the technical data sheet are not eligible for the defective media retention service feature option.

For more information To learn more on services for HP ESSN Options, please contact your HP sales representative or HP Authorized Channel Partner. Or visit: <http://www.hp.com/services/proliant> or www.hp.com/services/bladesystem

Related Options

HP High Performance Cluster Models

HP Insight Cluster Management Utility 1yr 24x7 Flexible License

QL803B

NOTE: This part number can be used to purchase one certificate for multiple licenses with a single activation key. Each license is for one node (server). Customer will receive a printed end user license agreement and license entitlement certificate via physical shipment. The license entitlement certificate must be redeemed online in order to obtain a license key.

NOTE: For additional license kits please see the QuickSpecs at:

http://h18004.www1.hp.com/products/quickspecs/12612_na/12612_na.html

HP Insight Cluster Management Utility 3yr 24x7 Flexible License

BD476A

NOTE: These part numbers can be used to purchase one certificate for multiple licenses and support with a single activation key. Each license is for one node (server). Customer will receive a printed end user license agreement and license entitlement certificate via physical shipment. The license entitlement certificate must be redeemed online in order to obtain a license key. Customer also will receive a support agreement.

HP Insight Cluster Management Utility Media

BD477A

NOTE: Order a minimum of one license per cluster to purchase media including software and documentation, which will be delivered to the customer, and also licenses CMU management. No license key is delivered or required

NOTE: For additional license kits please see the QuickSpecs at:

http://h18004.www1.hp.com/products/quickspecs/12612_na/12612_na.html

Technical Specifications

Form Factor	10.7 in (27.2 cm) PCIe x16 form factor	
Number of Tesla GPUs	Tesla K20, K20X	1 GPU
	Tesla K10	2 GPUs
Double Precision floating point performance (peak)	Tesla K10	190 Gflops (95 Gflops per GPU)
	Tesla K20	1.17 Tflops
	Tesla K20X	1.32 Tflops
Single Precision floating point performance (peak)	Tesla K10	4.577 Tflops (2.288 Tflops per GPU)
	Tesla K20	3.52 Tflops
	Tesla K20X	3.95 Tflops
Total Dedicated Memory	Tesla K10	8GB GDDR5 (4 GB per GPU)
	Tesla K20	5GB GDDR5
Memory Interface	384-bit	
Memory Bandwidth	Tesla M2090	178 GB/sec
	Tesla K10	320 GB/sec (160 GB per GPU)
	Tesla K20	200 GB/sec
	Tesla K20X	250 GB/sec
Power Consumption	Tesla K20	225W TDP
	Tesla K10, K20X	235W TDP
System Interface	Tesla K20	PCIe x16 Gen2
	Tesla K10, K20X	PCIe x16 Gen3
Thermal Solution	Passive heatsink cooled by host system airflow	

Environment-friendly Products and Approach	End-of-life Management and Recycling	Hewlett-Packard offers end-of-life HP product return, trade-in, and recycling programs in many geographic areas. For trade-in information, please go to: http://www.hp.com/go/green . To recycle your product, please go to: http://www.hp.com/go/green or contact your nearest HP sales office. Products returned to HP will be recycled, recovered or disposed of in a responsible manner.
---	---	--

The EU WEEE directive (2002/95/EC) requires manufacturers to provide treatment information for each product type for use by treatment facilities. This information (product disassembly instructions) is posted on the Hewlett Packard web site at: <http://www.hp.com/go/green>. These instructions may be used by recyclers and other WEEE treatment facilities as well as HP OEM customers who integrate and re-sell HP equipment.

Technical Specifications

© Copyright 2013 Hewlett-Packard Development Company, L.P.

The information contained herein is subject to change without notice.

Windows and Microsoft are registered trademarks of Microsoft Corp., in the U.S.

The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.